Journal of Multivariate Analysis **81**, 229–241 (2002)
doi:10.1006/jmva.2001.2004, available online at http://www.idealibrary.com on IDEAL®

Testing for Ordered Trends of Binary Responses between Contingency Tables

Chul Gyu Park

University of Suwon, Suwon, Korea

Received July 17, 1998; published online February 12, 2002

In this article, likelihood ratio tests (LRTs) are developed for detecting that stochastic trends of binary responses are ordered between $2 \times k$ contingency tables. We provide a simple iterative algorithm for the maximum likelihood estimators under the order restriction and construct the LRTs using those estimators. All the distributional results of these tests are based on the large sampling theory. The finite-sample behaviors of these tests are investigated through a simulation study. As an illustration of these tests, we analyze a set of data on wheeziness of smoking coalminers.

AMS 1991 subject classifications: 62F30; 62F10.

Key words and phrases: contingency tables; positive dependence; ordered trends; likelihood ratio test; chi-bar-square distribution.

1. INTRODUCTION

Stochastic dependence between ordered categorical variables has been a very important concept in the area of categorical data analysis. Suppose that X and Y are ordinal categorical variables. The relationship between X and Y can be described by virtue of the stochastic behavior of X according to change of Y. As Lehmann (1966) conceptualizes, there is positive (negative) dependence between X and Y if the conditional variable of X given $Y = y$ (denoted by $X(y)$) is stochastically increasing (decreasing) in y. If the conditional variable $X(y)$ is stochastically increasing until y reaches a point, and decreasing after that point of y, such a relationship is said to be unimodal dependence as discussed in Park (1998b).

Very often X might be a binary response variable while Y has k levels or k ordered categories. Then, the joint distribution of these variables is expressed as a $2 \times k$ contingency table with cell probabilities $p_{ij} = P[X = i, Y = j]$, $i = 1, 2$, $j = 1, 2, \ldots, k$. Let $\theta_{1j} = p_{2j}/(p_{1j} + p_{2j})$. Positive dependence between X and Y in this case is equivalent to the inequalities

$$\theta_{11} \leqslant \theta_{12} \leqslant \cdots \leqslant \theta_{1k} \tag{1.1}$$

0047-259X/02 $35.00

with at least one strict inequality. Of course, if all the inequalities in (1.1) hold as an equality, then X and Y are independent.

Testing procedures for independence between categorical variables are found in many papers. For $2\times k$ tables, Armitage (1955) proposed a test for independence against all possible linear trends between those variables. On the other hand, Grove (1980), Patefield (1982) and Lee (1989) considered positive dependence relationship as the alternative hypothesis. Estimation and testing issues on positive association in $R\times C$ tables have also been discussed in various association models. Recent works on this subject are Agresti *et al.* (1987), Gilula *et al.* (1988), Ritov and Gilula (1991), and Ritov and Gilula (1993), and many others. However, most of these works primarily aim at detecting the existence of positive dependence or association in a single contingency table with ordered categories.

Sometimes, one might be interested in comparing two categorical groups involved in a $2\times 2\times k$ table. Cochran (1955), Mantel and Haenszel (1959) and Mantel (1963) proposed non-model-based tests of conditional independence that apply to k strata of 2×2 tables. Trends of binary responses might also be compared between two independent groups each having $2\times k$ table. Standard procedure for this is to compare model coefficients involved in probit or logistic regression models. In this case there remains an argument that the scores are completely arbitrary. See Agresti (1990) for a brief summary of these models.

However, the comparison can be based on the concept of cone order discussed in Cohen and Sachrowitz (1996). Consider two $2\times k$ tables, $P=[p_{ij}]$ and $Q=[q_{ij}]$. Let $\tau_{1j}=q_{2j}/(q_{1j}+q_{2j})$ and $\delta_j=\theta_{1j}-\tau_{1j}$ for $j=1, 2, \ldots, k$. Also, let $\theta_1=(\theta_{11}, \theta_{12}, \ldots, \theta_{1k})'$ and $\tau_1=(\tau_{11}, \tau_{12}, \ldots, \tau_{1k})'$. If $\Delta=(\delta_1, \delta_2, \ldots, \delta_k)'\in I=\{x\in R^k : x_1\leqslant x_2\leqslant \cdots \leqslant x_k\}$, this implies that the parameter vector θ_1 locates more inside or closely to the isotonic cone of (1.1) than τ_1 because θ_1 in this case is the vector obtained by moving τ_1 to the direction of a vector in I. Thus, our comparison might be possible by testing

$$H_0\colon \theta_{11}-\tau_{11}=\theta_{12}-\tau_{12}=\cdots=\theta_{1k}-\tau_{1k} \tag{1.2}$$

against H_1-H_0 where the hypothesis H_1 is

$$H_1\colon \theta_{11}-\tau_{11}\leqslant\theta_{12}-\tau_{12}\leqslant\cdots\leqslant\theta_{1k}-\tau_{1k}. \tag{1.3}$$

For an illustration, consider smoking coalminers data discussed in Section 5. These data support in statistical sense that wheeziness and age are positively dependent in both groups of those workers with and without breathlessness. Now, we may wonder which of the two groups would have sharper trend of wheeziness rate against age. This question can be formulated as the above testing problem. Usual chi-square test for H_0 against all alternatives fails to reject H_0 even at 10% level in this example. However, H_0 is strongly rejected in our test of H_0 versus $H_1 - H_0$ with p-value being 0.023. Since H_1 is fairly supported by these data, our decision might be based on the test within the restricted parameter space.

In this context, we develop likelihood ratio tests for H_0 versus $H_1 - H_0$ and H_1 versus $H_2 - H_1$ where H_2 imposes no restriction. The latter test is a goodness-of-fit test that will be used for validating the order restriction in H_1. This restriction also coincides with the interpretation that the interaction effects between group and the levels of Y are ordered in the same direction. As we see in Section 4, the order restricted test is more powerful at some alternatives than a standard test based on logistic regression. Another advantage of our test is that it does not require determining scores which will affect the result of testing in standard approaches.

2. MAXIMUM LIKELIHOOD ESTIMATION

Consider two independent sets of cross-classified data whose cell frequencies are n_{ij} and m_{ij}, $i = 1, 2$, $j = 1, 2, \ldots, k$, and distributed with $[p_{ij}]$ and $[q_{ij}]$, respectively. Let $n_{+j} = n_{1j} + n_{2j}$ and $n_{++} = \sum_{j=1}^{k} n_{+j}$, and define m_{+j} and m_{++} similarly. With the additional reparameterization given by $\theta_{2j} = p_{1j} + p_{2j}$ and $\tau_{2j} = q_{1j} + q_{2j}$, $j = 1, 2, \ldots, k$, the likelihood function can be expressed as

$$L(\theta, \tau) = \prod_{j=1}^{k} [\{\theta_{1j}^{n_{2j}}(1-\theta_{1j})^{n_{1j}}\}\{\tau_{1j}^{m_{2j}}(1-\tau_{1j})^{m_{1j}}\}] \times \prod_{j=1}^{k} [\theta_{2j}^{n_{+j}}\tau_{2j}^{m_{+j}}]. \tag{2.1}$$

By maximizing the log-likelihood function, the unconstrained maximum likelihood estimators (MLEs) can be easily obtained, and they are

$$\hat{\theta}_{1j} = \frac{n_{2j}}{n_{+j}}, \ \hat{\theta}_{2j} = \frac{n_{+j}}{n_{++}}, \ \hat{\tau}_{1j} = \frac{m_{2j}}{m_{+j}}, \ \hat{\tau}_{2j} = \frac{m_{+j}}{m_{++}}, \qquad j = 1, 2, \ldots, k.$$

The estimation procedures for the MLEs under H_0 and H_1 are not simple. However, since those hypotheses do not impose any restriction on θ_{2j}'s and τ_{2j}'s, the estimators of those nuisance parameters under H_0 and H_1 (denoted by $\bar{\theta}_{2j}$, θ_{2j}^* and $\bar{\tau}_{2j}$, τ_{2j}^*, respectively) are obtained by maximizing the second product part of (2.1). Hence, they are exactly the same as the unconstrained estimators, and given by $\bar{\theta}_{2j} = \theta_{2j}^* = \hat{\theta}_{2j}$ and $\bar{\tau}_{2j} = \tau_{2j}^* = \hat{\tau}_{2j}$. Now, we need to find the constrained MLEs of θ_{1j}'s and τ_{1j}'s by maximizing the first product part of the likelihood function in (2.1) under H_0 and H_1, respectively. For this, let $\delta_j = \theta_{1j} - \tau_{1j}$, $j = 1, 2, \ldots, k$. Then, our hypotheses are expressed as

$$H_0\colon \delta_1 = \delta_2 = \cdots = \delta_k \qquad \text{and} \qquad H_1\colon \delta_1 \leqslant \delta_2 \leqslant \cdots \leqslant \delta_k.$$

Let $\Delta = (\delta_1, \delta_2, \ldots, \delta_k)'$ and $\tau_1 = (\tau_{11}, \tau_{12}, \ldots, \tau_{1k})'$. In this setup, the constrained MLEs are obtained by minimizing

$$H(\Delta, \tau_1) = -\sum_{j=1}^{k} [n_{2j} \ln(\tau_{1j} + \delta_j) + n_{1j} \ln(1 - \tau_{1j} - \delta_j) + m_{2j} \ln \tau_{1j} + m_{1j} \ln(1 - \tau_{1j})] \tag{2.2}$$

subject to the restrictions in H_0 and H_1, respectively.

In both cases, there is no explicit form of expressions for the constrained MLEs. However, we may suggest a simple iterative algorithm which requires solving only univariate nonlinear equations. For $j = 1, 2, \ldots, k$, let

$$S_j(\delta_j, \tau_{1j}) = \frac{\partial}{\partial \tau_{1j}} H(\Delta, \tau_1) = -\frac{n_{2j}}{\tau_{1j} + \delta_j} + \frac{n_{1j}}{1 - \tau_{1j} - \delta_j} - \frac{m_{2j}}{\tau_{1j}} + \frac{m_{1j}}{1 - \tau_{1j}}$$

and

$$R_j(\delta_j, \tau_{1j}) = \frac{\partial}{\partial \delta_j} H(\Delta, \tau_1) = -\frac{n_{2j}}{\tau_{1j} + \delta_j} + \frac{n_{1j}}{1 - \tau_{1j} - \delta_j}.$$

In order to obtain the H_0-constrained MLEs, we must solve for τ_{1j}'s and δ

$$S_j(\delta, \tau_{1j}) = 0, \qquad j = 1, 2, \ldots, k \tag{2.3}$$

and

$$\sum_{j=1}^{k} R_j(\delta, \tau_{1j}) = 0 \tag{2.4}$$

with constraints $0 < \tau_{1j} < 1$ and $0 < \tau_{1j} + \delta < 1$, $j = 1, 2, \ldots, k$. These solutions can be obtained by using the following algorithm:

Step 1. Set $\delta_\circ = 0$.

Step 2. For each j, solve $S_j(\delta_\circ, \tau_{1j}) = 0$ about τ_{1j} in the interval $[\max(0, -\delta_\circ), \min(1, 1 - \delta_\circ)]$, and then denote the solution by τ_{1j}°.

Step 3. Solve $\sum_{j=1}^{k} R_j(\delta, \tau_{1j}^\circ) = 0$ about δ in the interval $[-\min(\tau_{1j}^\circ), 1 - \max(\tau_{1j}^\circ)]$, and update $\delta_\circ$ by the new solution.

Step 4. Repeat Steps 2–3 until a satisfactory accuracy is achieved.

Since $S_j(\delta_\circ, \tau_{1j})$ in Step 2 and $\sum_{j=1}^{k} R_j(\delta, \tau_{1j}^\circ)$ in Step 3 are strictly monotone functions of τ_{1j} and δ, respectively, within the solution sets given in those steps, the solutions for the equations in Steps 2 and 3 are uniquely determined by usual numerical methods for univariate equations.

For the H_1-constrained MLEs, we must solve the problem of minimizing $H(\Delta, \tau_1)$ subject to

$$\delta_1 \leqslant \delta_2 \leqslant \cdots \leqslant \delta_k, \tag{2.5}$$

$$0 < \tau_{1j} < 1, \qquad 0 < \tau_{1j} + \delta_j < 1, \qquad j = 1, 2, \ldots, k. \tag{2.6}$$

Kuhn–Tucker conditions (see pp. 314–315 of Luenberger (1984)) for this minimization under the order restriction are equivalent to

$$S_j(\delta_j, \tau_{1j}) = 0, \qquad j = 1, 2, \ldots, k, \tag{2.7}$$

$$\sum_{l=1}^{j} R_l(\delta_l, \tau_{1l}) - \lambda_j = 0, \qquad j = 1, 2, \ldots, k-1, \qquad \sum_{l=1}^{k} R_l(\delta_l, \tau_{1l}) = 0, \tag{2.8}$$

$$\lambda_j(\delta_j - \delta_{j+1}) = 0, \qquad \lambda_j \geqslant 0, \qquad \delta_j - \delta_{j+1} \leqslant 0, \qquad j = 1, 2, \ldots, k-1, \tag{2.9}$$

where λ_j, $j = 1, 2, \ldots, k-1$ are Lagrangian multipliers. Thus, the solution satisfying (2.7), (2.8), (2.9) and (2.6) will be the desired estimators. These solutions, denoted by δ^* and τ_1^*, can be obtained by conducting repeatedly the computation procedures similar to those for H_0-estimators. Let $Av(i, j)$, $i \geqslant j$, be the solution of δ to

$$S_l(\delta, \tau_{1l}) = 0, \qquad l = i, i+1, \ldots, j, \tag{2.10}$$

$$\sum_{l=i}^{j} R_l(\delta, \tau_{1l}) = 0, \tag{2.11}$$

$$0 < \tau_{1l} < 1, \qquad 0 < \tau_{1l} + \delta < 1, \qquad l = i, i+1, \ldots, j. \tag{2.12}$$

Since these equations are the same types as those in (2.3) and (2.4), the solutions are attainable by the same algorithm used for H_0-constrained estimators. Similarly to the isotonization procedure in the case with quadratic objective function, the constrained MLE of δ_l is determined at $\delta_l^* = \max_{i \leqslant l} \min_{j \geqslant l} Av(i, j)$, $l = 1, 2, \ldots, k$. The proof is rather technical and will be omitted. For the H_1-constrained MLEs of τ_{1l}'s, let $\{i, i+1, \ldots, j\}$ be any level set in Δ^* such that $\delta_{i-1}^* < \delta_i^* = \cdots = \delta_j^* < \delta_{j+1}^*$. Then, τ_{1l}^*, $l = i, i+1, \ldots, j$ are the solutions to equations in (2.10) and are obtained when computing $Av(i, j)$.

3. TESTING PROCEDURES

Let $(\bar{\theta}, \bar{\tau})$, (θ^*, τ^*) and $(\hat{\theta}, \hat{\tau})$ be the MLE's of (θ, τ) under H_0, H_1 and H_2, respectively. Based upon these estimators, the likelihood ratio test (LRT) statistics for testing H_0 versus $H_1 - H_0$ and H_1 versus $H_2 - H_1$ are given by $T_{01} = 2[\ln L(\theta^*, \tau^*) - \ln L(\bar{\theta}, \bar{\tau})]$ and $T_{12} = 2[\ln L(\hat{\theta}, \hat{\tau}) - \ln L(\theta^*, \tau^*)]$, respectively. Thus, we reject H_0 and H_1 for the large values of T_{01} and T_{12}, respectively. Since $\bar{\theta}_{2j} = \theta^*_{2j} = \hat{\theta}_{2j}$ and $\bar{\tau}_{2j} = \tau^*_{2j} = \hat{\tau}_{2j}$ as discussed in Section 2, we can rewrite those LRT statistics as

$$T_{01} = 2[\ln L_1(\theta_1^*, \tau_1^*) - \ln L_1(\bar{\theta}_1 \bar{\tau}_1)] \tag{3.1}$$

and

$$T_{12} = 2[\ln L_1(\hat{\theta}_1, \hat{\tau}_1) - \ln L_1(\theta_1^*, \tau_1^*)], \tag{3.2}$$

where $L_1(\theta_1, \tau_1) = \prod_{j=1}^{k} [\{\theta_{1j}^{n_{2j}} (1-\theta_{1j})^{n_{1j}}\}\{\tau_{1j}^{m_{2j}}(1-\tau_{1j})^{m_{1j}}\}]$ with parameter vectors $\theta_1 = (\theta_{11}, \theta_{12}, \ldots, \theta_{1k})'$ and $\tau_1 = (\tau_{11}, \tau_{12}, \ldots, \tau_{1k})'$. Of course, the usual chi-square statistic for testing H_0 against $H_2 - H_0$ is given by

$$T_{02} = 2[\ln L_1(\hat{\theta}_1, \hat{\tau}_1) - \ln L_1(\bar{\theta}_1, \bar{\tau}_1)].$$

The null distributions of T_{01} and T_{12} are completely unknown in the finite-sample case. Hence, we appeal to asymptotic theories to determine their critical values. Suppose n_{++} and m_{++} go to infinity satisfying $\hat{\gamma} = m_{++}/n_{++} \to \gamma\ (>0)$. Following Lemma C in Serfling (1980, p. 154), we may approximate $\ln L_1(\hat{\theta}_1, \hat{\tau}_1) - \ln L_1(\theta_1, \tau_1)$ by

$$Q(\theta_1, \tau_1) = \tfrac{1}{2} n_{++} [(\hat{\theta}_1 - \theta_1)' I_{\theta_1} (\hat{\theta}_1 - \theta_1) + \hat{\gamma}(\hat{\tau}_1 - \tau_1)' I_{\tau_1} (\hat{\tau}_1 - \tau_1)] \tag{3.3}$$

with $o_p(n_{++}^{-1/2})$ error where I_{θ_1} and I_{τ_1} are information matrices for θ_1 and τ_1, respectively. Since $\partial^2/(\partial\theta_{1i}\,\partial\theta_{1j}) \ln L_1 = 0$ for $i \neq j$, the information matrix, I_{θ_1}, is a diagonal matrix. Let α_j be the jth element of I_{θ_1}. Then, since $p_{1j} = \theta_{2j}(1-\theta_{1j})$ and $p_{2j} = \theta_{1j}\theta_{2j}$, it can be easily verified that

$$\alpha_j = -\frac{1}{n_{++}} E\left[\frac{\partial^2}{\partial \theta_{1j}^2} \ln L_1\right] = \frac{\theta_{2j}}{\theta_{1j}(1-\theta_{1j})}, \qquad j = 1, 2, \ldots, k.$$

Similarly, we can show that I_{τ_1} is also a diagonal matrix with jth diagonal element

$$\beta_j = \frac{\tau_{2j}}{\tau_{1j}(1-\tau_{1j})}, \qquad j = 1, 2, \ldots, k.$$

These results imply that $\sqrt{n_{++}}\,(\hat{\theta}_{1j}-\theta_{1j})$, $j=1, 2, ..., k$ are asymptotically independent and normally distributed with their means 0 and variances α_j^{-1}, $j=1, 2, ..., k$. The same asymptotic properties are obtained for $\hat{\tau}_{1j}$, $j=1, 2, ..., k$. Using the facts discussed above, we may derive the following theorem.

THEOREM 1. *Let T_{01} and T_{12} be the LRT statistics for testing H_0 against H_1-H_0 and H_1 against H_2-H_1, respectively. Then, under H_0, we have for any $x \geqslant 0$*

$$\lim_{n_{++}\to\infty} P[T_{01} > x] = \sum_{j=1}^{k} p(j, k; w)\, P[\chi^2_{j-1} > x], \tag{3.4}$$

and

$$\lim_{n_{++}\to\infty} P[T_{12} > x] = \sum_{j=1}^{k} p(j, k; w)\, P[\chi^2_{k-j} > x], \tag{3.5}$$

where $w_j = \gamma\alpha_j\beta_j/(\alpha_j+\gamma\beta_j)$, $j=1, 2, ..., k$.

Proof. Let $\hat{\alpha}_j = \hat{\theta}_{2j}/(\hat{\theta}_{1j}(1-\hat{\theta}_{1j}))$ and $\hat{\beta}_j = \hat{\tau}_{2j}/(\hat{\tau}_{1j}(1-\hat{\tau}_{1j}))$. Since $\hat{\alpha}_j$ and $\hat{\beta}_j$ converges almost surely to α_j and β_j, respectively, $Q(\theta_1, \tau_1)$ converges almost surely to the same limit as

$$Q^*(\theta_1, \tau_1) = \tfrac{1}{2} n_{++} \left[\sum_{j=1}^{k} (\hat{\theta}_{1j}-\theta_{1j})^2\, \hat{\alpha}_j + \hat{\gamma} \sum_{j=1}^{k} (\hat{\tau}_{1j}-\tau_{1j})^2\, \hat{\beta}_j \right]. \tag{3.6}$$

Let the solution obtained by minimizing $Q^*(\theta_1, \tau_1)$ under H_0 and H_1 be denoted by $(\theta_1^\circ, \tau_1^\circ)$ and $(\tilde{\theta}_1, \tilde{\tau}_1)$, respectively. Also, let $\Delta^\circ = \theta_1^\circ - \tau_1^\circ$ and $\tilde{\Delta} = \tilde{\theta}_1 - \tilde{\tau}_1$. Then, as shown in Park (1998a), we have that

$$\tilde{\delta}_j = \max_{i\leqslant j}\, \min_{l\geqslant j}\, Sv(i, l), \qquad j=1, 2, ..., k \tag{3.7}$$

and

$$\tilde{\tau}_{1j} = \frac{1}{\hat{\alpha}_j + \hat{\gamma}\hat{\beta}_j} [\hat{\alpha}_j(\hat{\theta}_{1j}-\tilde{\delta}_j) + \hat{\gamma}\hat{\beta}_j\hat{\tau}_{1j}], \qquad j=1, 2, ..., k \tag{3.8}$$

where $Sv(i, j) = [\sum_{l=i}^{j} \hat{\gamma}\hat{\alpha}_l\hat{\beta}_l/(\hat{\alpha}_l+\hat{\gamma}\hat{\beta}_l)(\hat{\theta}_{1l}-\hat{\tau}_{1l})]/[\sum_{l=i}^{j} \hat{\gamma}\hat{\alpha}_l\hat{\beta}_l/(\hat{\alpha}_l+\hat{\gamma}\hat{\beta}_l)]$ for any (i, j) with $i \leqslant j$. The solution, $\tilde{\Delta}$, in (3.7) is the same as the isotonic regression of $\hat{\Delta} = \hat{\theta}_1 - \hat{\tau}_1$ with weights $\hat{w}_j = \hat{\gamma}\hat{\alpha}_j\hat{\beta}_j/(\hat{\alpha}_j + \hat{\gamma}\hat{\beta}_j)$, $j = 1, 2, \ldots, k$. It is also obvious that the H_0-constrained estimators are given by

$$\delta_1^\circ = \delta_2^\circ = \cdots = \delta_k^\circ = Sv(1, k) \tag{3.9}$$

and

$$\tau_{1j}^\circ = \frac{1}{\hat{\alpha}_j + \hat{\gamma}\hat{\beta}_j} [\hat{\alpha}_j(\hat{\theta}_{1j} - \delta_j^\circ) + \hat{\gamma}\hat{\beta}_j\hat{\tau}_{1j}], \qquad j = 1, 2, \ldots, k. \tag{3.10}$$

Putting $(\tilde{\theta}_1, \tilde{\tau}_1)$ and $(\theta_1^\circ, \tau_1^\circ)$ into (3.6), we can show that

$$Q^*(\tilde{\theta}_1, \tilde{\tau}_1) = \tfrac{1}{2} n_{++} \sum_{j=1}^{k} (\hat{\delta}_j - \tilde{\delta}_j)^2 \hat{w}_j$$

and

$$Q^*(\theta_1^\circ, \tau_1^\circ) = \tfrac{1}{2} n_{++} \sum_{j=1}^{k} (\hat{\delta}_j - \delta_j^\circ)^2 \hat{w}_j.$$

Using the properties of isotonic regression in Theorems 1.3.2 and 1.3.3 of Robertson *et al.* (1988), we can easily verify that

$$2[Q^*(\theta_1^\circ, \tau_1^\circ) - Q^*(\tilde{\theta}_1, \tilde{\tau}_1)] = n_{++} \sum_{j=1}^{k} (\tilde{\delta}_j - \delta_j^\circ)^2 \hat{w}_j \tag{3.11}$$

Now, recall that T_{01} and T_{12} are asymptotically the same as $2[Q^*(\theta_1^\circ, \tau_1^\circ) - Q^*(\tilde{\theta}_1, \tilde{\tau}_1)]$ and $2[Q^*(\tilde{\theta}_1, \tilde{\tau}_1)]$, respectively. Since $\sqrt{n_{++}}(\hat{\delta}_j - \delta_j)$, $j = 1, 2, \ldots, k$ are asymptotically independent and normally distributed with means 0 and variances w_j^{-1}, we have

$$T_{01} \xrightarrow{d} \sum_{j=1}^{k} (U_j^* - \bar{U}_j)^2 w_j \qquad \text{and} \qquad T_{12} \xrightarrow{d} \sum_{j=1}^{k} (U_j - U_j^*)^2 w_j$$

under H_0, where U^* is the isotonic regression of the random vector U following $N(0, W^{-1})$ with $W = diag\{w_1, w_2, \ldots, w_k\}$ and $\bar{U}_j = \sum_{l=1}^{k} w_l U_l / \sum_{l=1}^{k} w_l$, $j = 1, 2, \ldots, k$. Thus, the theorem is the immediate result of the Corollary of Theorem 2.3.1 of Robertson *et al.* (1988). ▮

In Theorem 1, $p(j, k; w)$ is the probability that the isotonic regression of the random vector Z following $N(0, W^{-1})$ takes on exactly j levels where $W = diag\{w_1, w_2, \ldots, w_k\}$. These level probabilities are generally unknown for $k \geqslant 5$. However, if there is no serious deviation among w_j's, for example, if $\max_j w_j/\min_j w_j \leqslant 1.4$, then one may use equal weights approximation which gives the recursive form

$$p(j,k)=\frac{1}{k}p(j-1,k-1)+\frac{k-1}{k}p(j,k-1) \quad \text{for} \quad j=2,3,\ldots,k-1 \quad (3.12)$$

with $p(1,k)=\frac{1}{k}$ and $p(k,k)=\frac{1}{k!}$. If those weights are seriously deviated, the level probabilities are usually estimated by simulating the distribution of the number of levels in the isotonic regression. See Chapter 2 of Robertson *et al.* (1988) for details. Since the weight vector w is unknown in the theorem, it is generally recommended to use the level probabilities obtained by estimating $p(j,k;\hat{w})$ where $\hat{w}_j=\hat{\gamma}\hat{\alpha}_j\hat{\beta}_j/(\hat{\alpha}_j+\hat{\gamma}\hat{\beta}_j)$, $j=1,2,\ldots,k$.

4. FINITE-SAMPLE PERFORMANCES

The null distributions of the test statistics developed in the previous section are based on the asymptotic results. In order to investigate their performances in the finite sample cases, a simulation study has been conducted for several underlying parameters in both null and alternative hypotheses. As a competitor, we considered logistic regression model formulated by

$$\log\frac{\pi_j}{1-\pi_j}=\alpha+\beta x_j,$$

where π_j is the conditional probability of success in the jth categorical level and $x_j=j$ are scores for ordered categories. The test to be compared with our test is the one-side test for the equality of slope parameters (β's) in two logistic models.

In this simulation study, we consider 2×4 contingency tables with nuisance parameters $\theta_{2j}=\tau_{2j}=0.25$, $j=1,2,3,4$. For the behaviors of the tests under the null hypothesis, we choose the parameters θ_{1j} and τ_{1j} such that $\delta_j=\theta_{1j}-\tau_{1j}=0.0$. Several parameter configurations for Δ are chosen for the alternative hypothesis. Sample sizes used in both groups are $(n_{++},m_{++})=(50,50),(100,100),(200,200)$.

Empirical sizes and powers of the order restricted test and logistic-model-based test were computed from 5000 replications, and they are listed in Table I. From this table, it is observed that both tests are rather conservative under small samples. However, with moderate sample sizes, say $(n_{++},m_{++})=(100,100)$, those tests are satisfactory in both size and power.

TABLE I

Empirical Sizes and Powers of Tests at Specified Nominal Levels

$\theta_{11}\theta_{12}\theta_{13}\theta_{14}$ $\tau_{11}\tau_{12}\tau_{13}\tau_{14}$	Sample size (n_{++}, m_{++})	T_{01} 5%	T_{01} 10%	T_{LG} [a] 5%	T_{LG} [a] 10%
		Null			
0.50 0.50 0.50 0.50	(50, 50)	0.038	0.079	0.034	0.078
0.50 0.50 0.50 0.50	(100, 100)	0.041	0.086	0.039	0.090
	(200, 200)	0.048	0.096	0.051	0.100
0.40 0.45 0.50 0.55	(50, 50)	0.033	0.070	0.033	0.079
0.40 0.45 0.50 0.55	(100, 100)	0.039	0.084	0.038	0.088
	(200, 200)	0.046	0.094	0.044	0.096
		Alternative			
0.40 0.60 0.60 0.60	(50, 50)	0.124	0.235	0.122	0.224
0.50 0.50 0.50 0.50	(100, 100)	0.233	0.356	0.217	0.349
	(200, 200)	0.403	0.541	0.362	0.512
0.40 0.40 0.60 0.60	(50, 50)	0.151	0.256	0.160	0.286
0.50 0.50 0.50 0.50	(100, 100)	0.283	0.425	0.310	0.450
	(200, 200)	0.529	0.667	0.542	0.679
0.40 0.40 0.40 0.60	(50, 50)	0.130	0.235	0.130	0.231
0.50 0.50 0.50 0.50	(100, 100)	0.230	0.356	0.215	0.336
	(200, 200)	0.413	0.553	0.370	0.518
0.40 0.45 0.50 0.55	(50, 50)	0.099	0.182	0.101	0.197
0.50 0.50 0.50 0.50	(100, 100)	0.154	0.259	0.167	0.284
	(200, 200)	0.255	0.384	0.279	0.418

[a] T_{LG} represents the test based on logistic regression.

For the parameter configurations in which δ_j are linearly increasing, logistic regression approach is more powerful than the order restricted test. However, if δ_j are not linearly increasing (for example, $\delta_1 = \delta_2 = \delta_3 < \delta_4$), our test achieves higher powers than the standard approach. Here, it should be noted that powers are affected by the choice of scores in the case of the test by logistic regression. Thus, this conventional test might be too good for some alternatives and too bad for some other alternatives, depending upon the score vector. From this point, the order restricted test might be more preferred because we need no score values for ordered categories.

5. AN EXAMPLE

For illustrating the testing methods developed in Section 3, we analyze the coalminers data originally reported in Table 1 of Ashford and Sowden (1970). These data are composed of the numbers of subjects cross-classified

TABLE II

Coalminers Data Classified by Breathlessness, Wheeze and Age

Age	Non-breathless		Breathless	
	No Wheeze	Wheeze	No Wheeze	Wheeze
20–24	1841	95	7	9
25–29	1654	105	9	23
30–34	1863	177	19	54
35–39	2357	257	48	121
40–44	1778	273	54	169
45–49	1712	324	88	269
50–54	1324	245	117	404
55–59	967	225	152	406
60–64	526	132	106	372

according to breathlessness, wheeziness, and age. They are tabulated in Table II.

In this analysis, we compare the trends of incidence rates of wheeziness over ages between the two groups specified by the state of breathlessness. Let θ_{1j} denote the incidence rates of wheeziness in the jth age group of subjects with no breathlessness. Similarly, the conditional incidence rates for the breathless subject groups are denoted by τ_{1j}, $j = 1, 2, \ldots, k$.

The estimates of parameters of our primary interests are listed in Table III. As we see from the unrestricted estimates, there is a fairly strong sign of positive dependence between wheeziness and age in both groups. This argument is clearly validated by the likelihood ratio tests of independence

TABLE III

Estimates of Parameters

Age group	Under H_2			Under H_1			Under H_0			
	$\hat{\theta}_1$	$\hat{\tau}_1$	$\hat{\Delta}$	θ_1^*	τ_1^*	Δ^*	$\bar{\theta}_1$	$\bar{\tau}_1$	$\bar{\Delta}$	$\hat{w}$
20–24	0.049	0.563	−0.513	0.049	0.686	−0.637	0.049	0.641	−0.592	0.004
25–29	0.060	0.719	−0.659	0.060	0.697	−0.637	0.060	0.652	−0.592	0.010
30–34	0.087	0.740	−0.653	0.087	0.724	−0.637	0.088	0.679	−0.592	0.024
35–39	0.098	0.716	−0.618	0.098	0.720	−0.622	0.099	0.691	−0.592	0.051
40–44	0.133	0.758	−0.625	0.133	0.755	−0.622	0.135	0.727	−0.592	0.072
45–49	0.159	0.754	−0.594	0.158	0.766	−0.609	0.159	0.751	−0.592	0.108
50–54	0.156	0.775	−0.619	0.158	0.767	−0.609	0.161	0.763	−0.592	0.151
55–59	0.189	0.728	−0.539	0.184	0.740	−0.555	0.174	0.766	−0.592	0.130
60–64	0.201	0.778	−0.578	0.209	0.765	−0.555	0.195	0.787	−0.592	0.104

versus positive dependence whose test statistics are computed as 350.2 and 5.762 with p-values 0.000 and 0.089 respectively in both groups. Now, the question will be which of the two groups shows sharper trend of incidence rates against age. First, note that the goodness-of-fit test based on test statistic T_{12} ($=3.257$) fairly supports the hypothesis $H_1: \delta_1 \leqslant \delta_2 \leqslant \cdots \leqslant \delta_k$ with p-value of 0.727. This fact also supports the use of test statistic T_{01} in testing $H_0: \delta_1 = \delta_2 = \cdots = \delta_k$. Since we have $T_{01} = 9.031$ with its p-value being 0.023, H_0 is rejected in favor of $H_1 - H_0$ at usual 5 or 10% levels. However, note that the usual chi-square test based on T_{02} ($=12.288$) fails to reject H_0 (p-value$=0.139$). From this analysis, it may be concluded that the group of miners without breathlessness have sharper trend of the incidence rate of wheeziness against age levels.

ACKNOWLEDGMENTS

The author thanks the referee and the editor for their helpful comments and recommendations. This research was supported by Grant 981-0105-030-1 from the Basic Research Program of the Korea Science and Engineering Foundation.

REFERENCES

A. Agresti, "Categorical Data Analysis," Wiley, New York, 1990.

A. Agresti, C. Chuang, and A. Kezouh, Order-restricted score parameters in association models for contingency tables, *J. Amer. Statist. Assoc.* **82** (1987), 619–623.

P. Armitage, Tests for linear trends in proportions and frequencies, *Biometrics* **11** (1955), 375–386.

J. R. Ashford and R. D. Sowden, Multivariate probit analysis, *Biometrics* **26** (1970), 535–546.

W. G. Cochran, A test of a linear function of the deviations between observed and expected numbers, *J. Amer. Statist. Assoc.* **50** (1955), 377–397.

A. Cohen and H. B. Sackrowitz, Cone order association and stochastic cone ordering with applications to order-restricted testing, *Ann. Statist.* **24** (1996), 2036–2048.

Z. Gilula, A. M. Krieger, and Y. Ritov, Ordinal association in contingency tables: some interpretive aspects, *J. Amer. Statist. Assoc.* **83** (1988), 540–545.

D. M. Grove, A test of independence against a class of ordered alternatives in a $2 \times C$ contingency table, *J. Amer. Statist. Assoc.* **75** (1980), 454–459.

M. T. Lee, Some cross-product difference statistics and a test for trends in ordered contingency tables, *Statist. Probab. Lett.* **7** (1989), 41–46.

E. L. Lehmann, Some concepts of dependence, *Ann. Math. Statist.* **26** (1966), 1137–1153.

D. G. Luenberger, "Linear and Nonlinear Programming," Addison–Wesley, Reading, MA, 1984.

N. Mantel, Chi-square tests with one degree of freedom: extensions of the Mantel–Haenszel procedure, *J. Amer. Statist. Assoc.* **58** (1963), 690–700.

N. Mantel and W. Haenszel, Statistical aspects of the analysis of data from retrospective studies of disease, *J. Natl. Cancer Inst.* **22** (1959), 719–748.

C. G. Park, Least squares estimation of two functions under order restriction in isotonicity, *Statist. Probab. Lett.* **37** (1998a), 97–100.

C. G. Park, Testing for unimodal dependence in an ordered contingency table with restricted marginal probabilities, *Statist. Probab. Lett.* **37** (1998b), 121–129.

W. M. Patefield, Exact tests for trends in ordered contingency tables, *Appl. Statist.* **31** (1982), 32–43.

Y. Ritov and Z. Gilula, The order-restricted RC model for ordered contingency tables: estimation and testing for fit, *Ann. Statist.* **19** (1991), 2090–2101.

Y. Ritov and Z. Gilula, Analysis of contingency tables by correspondence models subject to order constraints, *J. Amer. Statist. Assoc.* **88** (1993), 1380–1387.

T. Robertson, F. T. Wright, and R. L. Dykstra, "Order Restricted Statistical Inference," Wiley, New York, 1988.

R. J. Serfling, "Approximation Theorems of Mathematical Statistics," Wiley, New York, 1980.